



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/751,016	12/31/2003	Bhashyam Ramesh	11303	3665
7590	10/03/2006			EXAMINER SAEED, USMAAN
John D. Cowart Teradata Law IP, WHQ-4W NCR Corporation 1700 S. Patterson Blvd. Dayton, OH 45479-0001			ART UNIT 2166	PAPER NUMBER

DATE MAILED: 10/03/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)
	10/751,016	RAMESH ET AL.
	Examiner Usmaan Saeed	Art Unit 2166

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --
Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) Responsive to communication(s) filed on 31 December 2003.
 2a) This action is FINAL. 2b) This action is non-final.
 3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) Claim(s) 1-42 is/are pending in the application.
 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
 5) Claim(s) _____ is/are allowed.
 6) Claim(s) 1-42 is/are rejected.
 7) Claim(s) _____ is/are objected to.
 8) Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) The specification is objected to by the Examiner.
 10) The drawing(s) filed on 31 December 2003 is/are: a) accepted or b) objected to by the Examiner.
 Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
 Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
 11) The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
 a) All b) Some * c) None of:
 1. Certified copies of the priority documents have been received.
 2. Certified copies of the priority documents have been received in Application No. _____.
 3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|--|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413) |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | Paper No(s)/Mail Date. _____ |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
Paper No(s)/Mail Date _____ | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152) |
| | 6) <input type="checkbox"/> Other: _____ |

DETAILED ACTION

1. Claims 1-42 are pending in this office action.

Claim Rejections - 35 USC § 112

2. The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

Claims 4, 7, 18, 21, 32, and 35 are rejected under 35 U.S.C. 112 because they recite "(1+S), where S is a positive percentage represented as a decimal." It is unclear that S is positive percentage for what kind of values, either for minimum percentage of rows or for any other value. Appropriate correction is required.

Claim Objections

3. Claim 9, 23, and 37 objected to because of the following informalities: the limitation "determining a reminder number of buckets equal to the total number of buckets less the number of high-bias buckets used" needs grammatical revision. Appropriate correction is required.

Claim Rejections - 35 USC § 101

4. 35 U.S.C. 101 reads as follows:

Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

Claims 1-42 are rejected under 35 U.S.C. 101 as being directed to non-statutory subject matter. The language of the claims raises a question as to whether the claims are directed merely to an environment or machine which would result in a practical application producing a concrete useful, and tangible result to form the basis of statutory subject matter under 35 U.S.C. 101.

Claims 1-42 are rejected because the claims do not recite a practical application by producing a physical transformation or producing a useful, concrete, and tangible results. To perform a physical transformation, the claimed invention must transform an article of physical object into a different state or thing. Transformation of data is not a physical transformation. A useful, concrete, and tangible results must be either specifically recited in the claim or flow inherently therefrom. To be useful the claimed invention must establish a specific, substantial, and credible utility. To be concrete the claimed invention must be able to produce reproducible results. To be tangible the claimed invention must produce must produce a practical application or real world result.

To expedite a complete examination of the instant application the claims rejected under U.S.C. 101 (nonstatutory) above are further rejected as set forth

below in anticipation of application amending these claims to place them within the four categories of invention.

Claim Rejections - 35 USC § 102

5. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

Claims 1-2, 14-16, 28-30, and 42 are rejected under 35 U.S.C. 102(b) as being anticipated by **Kuorong Chiang (Chiang hereinafter)** (U.S. Patent No. 6,477,523).

With respect to claim 1, Chiang teaches “**a method for representing statistics about a table including one or more rows, each row including a respective value, the method including**” as an article of manufacture for generating statistics for use by a relational database management system (**Chiang Abstract**).

“**creating zero or more histogram buckets, each histogram bucket including a width representing a respective range of values and a height representing a count of rows having values in the range of values**” as in the

preferred embodiment, data partitioning and repartitioning may be performed, in order to enhance parallel processing across multiple AMPs 116. For example, the data may be hash partitioned, range partitioned, or not partitioned at all (i.e., locally processed) (**Chiang** Col 5, Lines 25-39). Wherein the ModeFreq field in the equal-heights interval represents a number of rows having a modal value (**Chiang** Col 10, Lines 20-22).

“creating one or more high-bias buckets, each high-bias bucket representing one or more values that appear in a minimum percentage of rows” as the compressed histogram includes both equal-height intervals and high-biased intervals (**Chiang** Abstract). Count of rows is stored in ModeFreq for the first Loner and is stored in the rows field for the second loner. Loner is a distinct values that is stored in a high-biased interval (**Chiang** Col 4, Lines 6-10). Examiner interprets loner values as having minimum percentage of rows, which are stored in high biased interval.

With respect to claim 2, **Chiang** teaches, **“a total number of buckets is a fixed number equal to the sum of the number of histogram buckets and the number of high-bias buckets”** as the compressed histogram includes both equal-height intervals and high-biased intervals (**Chiang** Abstract).

With respect to claim 14, **Chiang** discloses **the method of claim 1, where a total number of buckets is equal to the sum of a number of the histogram buckets and a number of the high-bias buckets, where the total**

number of buckets is fixed, where the number of high-bias buckets is fixed, and where the method includes: as the compressed histogram includes both equal-height intervals and high-biased intervals (**Chiang Abstract**).

“populating the one or more high-bias buckets with the FH most frequently occurring values, where F is a number of values each high-bias bucket can store and H is the number of high-bias buckets; and populating the one or more histogram buckets with all other values” as the compressed histogram includes both equal-height intervals and high-biased intervals (**Chiang Abstract**). The Values field represents the number of loners in the interval (**Chiang Col 9, Lines 66-67**). Compressed histogram is an array of intervals, which comprises high-biased or equal-height intervals, or both. In the latter situation, high-biased intervals are ordered before the equal-height intervals (**Chiang Col 4, Lines 17-20**).

Claims 15-16, 28-30, and 42 are essentially the same as claims 1, 2, and 14 except they set forth the claimed invention as a system and a computer program and are rejected for the same reasons as applied hereinabove.

Claim Rejections - 35 USC § 103

6. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

- (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to

Art Unit: 2166

be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

This application currently names joint inventors. In considering patentability of the claims under 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of 35 U.S.C. 103(c) and potential 35 U.S.C. 102(e), (f) or (g) prior art under 35 U.S.C. 103(a).

Claims 3-9, 17-23 and 31-37 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Kuorong Chiang**. (**Chiang** hereinafter) (U.S. Patent No 6,477,523) as applied to claims 1-2, 14-16, 28-30, and 42 in view of **Campos et al.** (**Campos** hereinafter) (U.S PG Pub No. 2003/0212702).

With respect to claim 3, **Chiang** teaches **the method of claim 1, where creating the high-bias and histogram buckets includes:**

“(a) determining an average height of the histogram buckets” as Global Interval Size--the average number of rows to be fitted in one interval (Chiang** Col 4, Lines 17-20).**

"(b) based on the average height of the histogram buckets, determining a reclassification threshold, (c) representing each value that exceeds the reclassification threshold in a high-bias bucket" as high-biased intervals store explicit column values and frequencies, so that a 100% estimation accuracy is obtained for these loners. Moreover, the rest of the column values can be made more uniform, if the column values with highest frequencies are removed from the equal-height intervals and put into high-biased ones. This way, not only do loners receive perfect estimation, but non-loners also benefit from increased uniformity (**Chiang** Col 2, Lines 12-20). Therefore the values with the highest frequencies are placed into the high biased buckets.

Chiang discloses the elements of claim 3 as noted above but does not explicitly teaches "**reclassification threshold.**"

However, **Campos** discloses "**reclassification threshold**" as when the number of entries assigned to a node reaches a pre-specified threshold the node is split and its buffer entries divided among its child nodes (**Campos** Paragraph 0052).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teaching of the cited references because **Campos's** teachings would have allowed **Chiang** to provides improved performance in model building and data mining, good integration with the various databases throughout the enterprise, and flexible specification and adjustment of the models being built, and which provides reductions in development times and costs for data mining projects (**Campos** Paragraph 0007).

With respect to claim 4, **Chiang** does not explicitly discloses “**the reclassification threshold is equal to the average height of the histogram bucket multiplied by (1+S), where S is a positive percentage represented as a decimal.**”

However, **Campos** discloses “**the reclassification threshold is equal to the average height of the histogram bucket multiplied by (1+S), where S is a positive percentage represented as a decimal**” as in step 1312, the average histogram height is computed for the non-zero bins $H=H_s/B$ where B is the number of non-zero bins and H_s is the sum of the heights for the non-zero bins (**Campos** Paragraph 0184). For each bin, if the bin height H_b is above a pre-defined small threshold (e.g., 10E-100), then $P_c=\max(\ln(H_b/H_p)+k)$ where P_c is the log conditional probability, and the constant k is used to make it compatible with the Nave Bayes implementation (**Campos** Paragraph 0187).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teaching of the cited references because **Campos's** teachings would have allowed **Chiang** to provides improved performance in model building and data mining, good integration with the various databases throughout the enterprise, and flexible specification and adjustment of the models being built, and which provides reductions in development times and costs for data mining projects (**Campos** Paragraph 0007).

With respect to claim 5, **Chiang** teaches “**the method of claim 3 where (a), (b), and (c) are repeated until no values exceeds the reclassification threshold**” as high-biased intervals store explicit column values and frequencies, so that a 100% estimation accuracy is obtained for these loners. Moreover, the rest of the column values can be made more uniform, if the column values with highest frequencies are removed from the equal-height intervals and put into high-biased ones. This way, not only do loners receive perfect estimation, but non-loners also benefit from increased uniformity (**Chiang** Col 2, Lines 12-20). All the values with the highest frequencies are removed from the equal-height intervals and put into high-biased ones.

Chiang discloses the elements of claim 5 as noted above but does not explicitly teaches “**reclassification threshold**.”

However, **Campos** discloses “**reclassification threshold**” as when the number of entries assigned to a node reaches a pre-specified threshold the node is split and its buffer entries divided among its child nodes (**Campos** Paragraph 0052).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teaching of the cited references because **Campos's** teachings would have allowed **Chiang** to provides improved performance in model building and data mining, good integration with the various databases throughout the enterprise, and flexible specification and adjustment of the models being built, and which provides reductions in development times and costs for data mining projects (**Campos** Paragraph 0007).

With respect to claim 6, Chiang teaches the method of claim 1, where creating the high-bias and histogram buckets includes:

"(a) determining an average height of the histogram buckets" as Global Interval Size--the average number of rows to be fitted in one interval (Chiang Col 4, Lines 17-20).

"(b) based on the average height of the histogram buckets, determining a reclassification threshold" as high-biased intervals store explicit column values and frequencies, so that a 100% estimation accuracy is obtained for these loners. Moreover, the rest of the column values can be made more uniform, if the column values with highest frequencies are removed from the equal-height intervals and put into high-biased ones. This way, not only do loners receive perfect estimation, but non-loners also benefit from increased uniformity (Chiang Col 2, Lines 12-20).

"(c) for each value that exceeds the reclassification threshold:

(1) if all of the high-bias buckets are not full, representing the value in a high-bias bucket" as high-biased intervals store explicit column values and frequencies, so that a 100% estimation accuracy is obtained for these loners. Moreover, the rest of the column values can be made more uniform, if the column values with highest frequencies are removed from the equal-height intervals and put into high-biased ones. This way, not only do loners receive perfect estimation, but non-loners also benefit from increased uniformity (Chiang Col 2, Lines 12-20).

Chiang teaches the elements of claim 6 as noted above but does not explicitly discloses “**reclassification threshold**” and “**(2) else, if the number of high-bias buckets is less than a fixed number of high-bias buckets:**

- (i) creating a new high-bias bucket; and**
- (ii) representing the value in the new high-bias bucket.”**

However, **Campos** discloses “**reclassification threshold**” and “**(2) else, if the number of high-bias buckets is less than a fixed number of high-bias buckets: (i) creating a new high-bias bucket; and**

(ii) representing the value in the new high-bias bucket” as when the number of entries assigned to a node reaches a pre-specified threshold the node is split and its buffer entries divided among its child nodes (**Campos** Paragraph 0052).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teaching of the cited references because **Campos’s** teachings would have allowed **Chiang** to provides improved performance in model building and data mining, good integration with the various databases throughout the enterprise, and flexible specification and adjustment of the models being built, and which provides reductions in development times and costs for data mining projects (**Campos** Paragraph 0007).

Claims 7 and 8 are same as claims 4 and 5 and are rejected for the same reasons as applied hereinabove.

With respect to claim 9, Chiang teaches the method of claim 1, where a total number of buckets is equal to the sum of a number of histogram buckets and a number of high-bias buckets, where the total number of buckets is fixed, and where the method further includes:

"(a) identifying one or more values that appear in at least the minimum percentage of rows and representing the identified values in the high-bias buckets" as the compressed histogram includes both equal-height intervals and high-biased intervals (Chiang Abstract). Count of rows is stored in ModeFreq for the first Loner and is stored in the rows field for the second loner. Loner is a distinct values that is stored in a high-biased interval (Chiang Col 4, Lines 6-10).

"(b) determining a remaining number of buckets equal to the total number of buckets less the number of high-bias buckets used" as if, at anytime, the count of a row of the global aggregate spool is greater than or equal to the Loner criteria, then the summary record's count field is set to (-1)*(row's count) and the summary record is sent to the coordinator AMP 116 (Chiang Col 7, Lines 14-18).

"(c) if the number of remaining buckets is greater than a stop number of buckets: (1) adjusting the minimum percentage of rows; (2) identifying values that appear in the adjusted minimum percentage of rows; and (3) representing values that appear in the adjusted minimum percentage of row in high-bias buckets" as the compressed histogram includes both equal-height intervals and high-biased intervals (Chiang Abstract).

Count of rows is stored in ModeFreq for the first Loner and is stored in the rows field for the second loner. Loner is a distinct values that is stored in a high-biased interval (**Chiang** Col 4, Lines 6-10). Examiner interprets loner values as having minimum percentage of rows, which are stored in high biased interval.

Claims 17-23 and 31-37 are essentially the same as claims 3-9 except they set forth the claimed invention as a system and a computer program and are rejected for the same reasons as applied hereinabove.

7. Claims 10-13, 24-27 and 38-41 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Kuorong Chiang**. (**Chiang** hereinafter) (U.S. Patent No 6,477,523) in view of **Campos et al.** (**Campos** hereinafter) (U.S PG Pub No. 2003/0212702) as applied to claims 3-9, 17-23 and 31-37 further in view of **Ari W. Mozes** (**Mozes** hereinafter) (U.S. Patent No 6,691,099).

With respect to claim 10 and 11, **Chiang and Campos** do not explicitly teach “**the method of claim 9, where (a) includes setting the minimum percentage of rows to 1/(FB)% where F is equal to a number of high-bias values that each high-bias bucket can contain and B is equal to the total number of buckets**” and the method of claim 9, where **(c)(1) includes setting the adjusted minimum percentage to (V(FB - 1))/ FB %, where F is equal to a number of high-bias values that each high-bias bucket can contain, B is equal to the total number of buckets, V is equal to the**

minimum percentage of rows, and I is equal to a number of values represented in high-bias buckets."

However, **Mozes** discloses "the method of claim 9, where (a) includes setting the minimum percentage of rows to $1/(FB)\%$ where F is equal to a number of high-bias values that each high-bias bucket can contain and B is equal to the total number of buckets" and the method of claim 9, where (c)(1) includes setting the adjusted minimum percentage to $(V(FB - 1))/ FB\%$, where F is equal to a number of high-bias values that each high-bias bucket can contain, B is equal to the total number of buckets, V is equal to the minimum percentage of rows, and I is equal to a number of values represented in high-bias buckets" as for example, consider if the statistic being addressed by the sampling is the "Number of Rows in Table." A minimum value, such as "2500" can be established for this type of statistic. If the identified number of rows from step 202 is less than 2500 rows, then the sample size or sample percentage is increased (208), and steps 202 and 204 are repeated until the minimum sample size is achieved (**Mozes** Col 4, Lines 47-54).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teaching of the cited references because **Mozes's** teachings would have allowed **Chiang and Campos** to provide a mechanism for automatically determining an adequate sample size for both statistics and histograms (**Mozes** Col 3, Lines 27-35).

With respect to claim 12, **Chiang teaches the method of claim 9, further including:**

"(d) if the number of remaining buckets is less than or equal to the stop number of buckets: representing values not represented in high-bias buckets in histogram buckets" as the compressed histogram includes both equal-height intervals and high-biased intervals (**Chiang Abstract**).

Chiang teaches the elements of claim 12 as noted above but does not explicitly discloses "**the number of remaining buckets is less than or equal to the stop number of buckets.**"

However, **Mozes** discloses "**the number of remaining buckets is less than or equal to the stop number of buckets**" as the sampling rate is adjusted upward to collect an adequate sample size. In one embodiment, if the number of non-null column values in the sample is less than 2500, then the sample rate is increased to provide more samples (**Mozes** Col 5, Lines 31-35).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teaching of the cited references because **Mozes's** teachings would have allowed **Chiang and Campos** to provides a mechanism for automatically determining an adequate sample size for both statistics and histograms (**Mozes** Col 3, Lines 27-35).

With respect to claim 13, **Chiang and Campos** do not explicitly disclose "**(e) repeating (b), (c), and (d) until the number of remaining buckets is less than or equal to the stop number of buckets.**"

However, **Mozes** discloses ““(e) repeating (b), (c), and (d) until the number of remaining buckets is less than or equal to the stop number of buckets” as the sampling rate is adjusted upward to collect an adequate sample size. In one embodiment, if the number of non-null column values in the sample is less than 2500, then the sample rate is increased to provide more samples (**Mozes** Col 5, Lines 31-35).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teaching of the cited references because **Mozes**'s teachings would have allowed **Chiang and Campos** to provides a mechanism for automatically determining an adequate sample size for both statistics and histograms (**Mozes** Col 3, Lines 27-35).

Claims 24-27 and 38-41 are essentially the same as claims 10-13 except they set forth the claimed invention as a system and a computer program and are rejected for the same reasons as applied hereinabove.

Conclusion

8. The prior art made of record and not replied upon is considered pertinent to applicant's disclosure is listed on 892 form.

Contact Information

Art Unit: 2166

9. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Usmaan Saeed whose telephone number is (571)272-4046. The examiner can normally be reached on M-F 8-5.

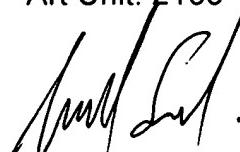
If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Hosain Alam can be reached on (571)272-3978. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).



Leslie Wong
Primary Examiner

Usmaan Saeed
Patent Examiner
Art Unit: 2166



US
September 25, 2006